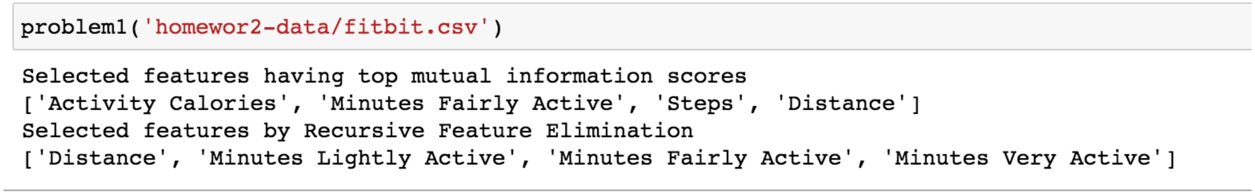
BBL536E Homework-2 Report

Problem-1.

In this task, I have tried to find top 4 features , most related with the target using Mutual Information score and Recursive Feature Elimination method .

Outputs for these two methods are as follows;

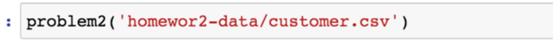


```
problem1('homewor2-data/fitbit.csv')
```

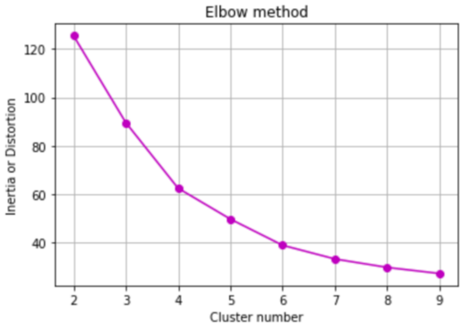
```
Selected features having top mutual information scores
['Activity Calories', 'Minutes Fairly Active', 'Steps', 'Distance']
Selected features by Recursive Feature Elimination
['Distance', 'Minutes Lightly Active', 'Minutes Fairly Active', 'Minutes Very Active']
```

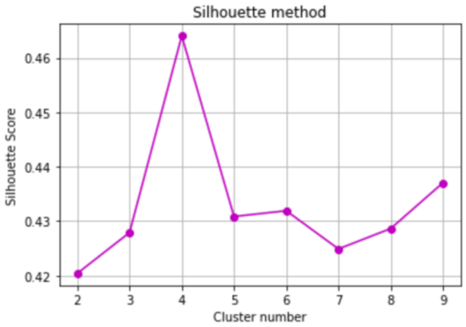
Problem2.

Using K-Means algorithm with scaled features, I have produced elbow and silhouette coefficients as follows;



```
: problem2('homewor2-data/customer.csv')
```





Looking at the elbow graph, best cluster number might be 6 or 7, since after these cluster numbers, effect of reducing distorsion is not that much.

For silhouette graph, cluster number might be 4, since cluster number 4 gives the highest score and closer to 1 (best score) compared to the others.

Problem-3.

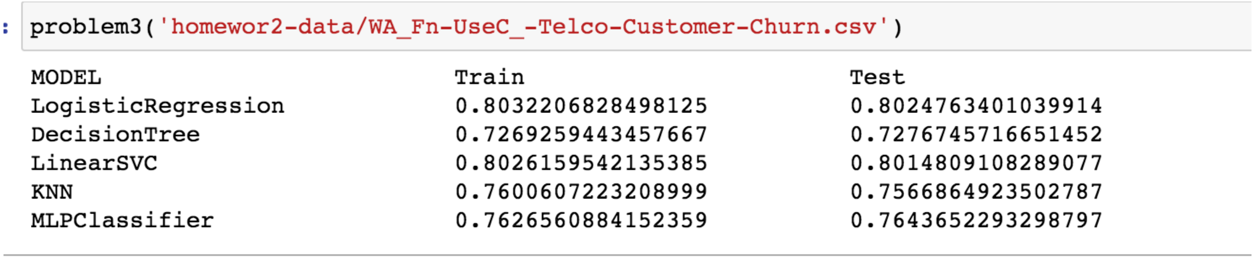
In problem 3, i used scaled features.

For each model, K-Fold cross validation with shuffling strategy was used to split the data.

In each fold, i have trained models with training set and calculated the average cross validation accuracy scores to validate power of my models.

Then accuracy scores of our models on test splits were also calculated.

Obtained outputs are as follows;



```
problem3('homewor2-data/WA_Fn-UseC_-Telco-Customer-Churn.csv')
```

```
MODEL                         Train                         Test
LogisticRegression            0.8032206828498125            0.8024763401039914
DecisionTree                  0.7269259443457667            0.7276745716651452
LinearSVC                     0.8026159542135385            0.8014809108289077
KNN                           0.7600607223208999            0.7566864923502787
MLPClassifier                 0.7626560884152359            0.7643652293298797
```